

Simulating the Evolution of a Grammar for Case

Luc Steels
Sony CSL Paris
VUB AI lab - Brussels
steels@arti.vub.ac.be

Abstract

The paper presents a multi-agent model in which steps towards the evolution of a grammar for case are simulated. The agents have the capability to visually perceive dynamic scenes and analyse them into event structures. They then play a language game in which one agent describes to another one a recently seen event. Agents invent and learn grammar as part of the game. Grammar is seen as an evolving adaptive system that is shaped and negotiated by the agents as they optimise communication.

Keywords: case grammar, evolutionary linguistics, multi-agent simulations.

1 Introduction

This paper reports on research attempting to simulate the emergence and evolution of grammar through multi-agent simulations. The agents are embodied and situated in the world and the language is grounded in reality through the sensory-motor apparatus of the agents [19]. In each simulation, a population of agents is set up, with each agent having a particular cognitive architecture. The agents engage in repeated interactions, called language games. Artificial lexicons and grammars emerge as agents invent and learn grammar in order to be (more) successful in the game.

The goal of these simulations is to determine precisely what kind of cognitive structures and interactive behaviors need to be introduced for particular language phenomena to emerge. Of course, the artificial agents will never evolve specific natural languages, like English or Swahili, simply because many factors influence the shape of a language and it is impossible to simulate all of them. Moreover there are historical contingencies. The point is rather to try and evolve an artificial language that has those characteristics of human language one is trying to investigate.

Several other examples of such a multi-agent modeling approach have been developed recently [3] [5]. One aspect, which has been studied quite successfully, is the emergence of compositionality through iterated learning [11]. Also for the emergence and evolution of lexicons, very solid results have been achieved, even in large-scale populations that have to deal with an open-ended environment [22]. This paper studies

whether we can evolve grammars that are closer to natural grammars than those obtained so far. The mechanisms we will use have been heavily inspired by research on grammaticalisation and its underlying cognitive mechanisms [9].

From a functional or cognitive grammar viewpoint [13], grammar expresses aspects of meaning using form ingredients like word order, affixes, inflection, function words or intonation. Semantic domains that are typically grammaticalised include event-object relations (as in grammars for case [1]), tense - aspect - mood - modality, determination, propositional structure, information structure, reflexivity, etc. But there are significant differences between languages with respect to whether they will grammaticalise a particular domain or not (some languages do not grammatically express tense for example) and also how the domain is expressed in a particular language.

The experiments reported here focus only on one specific aspect of grammar, typically expressed with a grammar of case. Grammars of case are primarily concerned with the relation between the descriptions of events (typically by a verb) and the objects participating in the event. For example, a dative noun in Latin as used with the verb “donnare” (give) expresses that the entity referred to by the noun is the recipient. Languages like English which have a very weak case system, rely heavily on word order and prepositions for the same purpose. When prepositions have become function words (like the preposition ‘by’ in English) they behave similar to case markers or particles.

The experiments reported in this paper adopt the hypothesis that grammar invention and grammar learning is driven by the desire to communicate as effectively as possible. This means that speakers try to avoid ambiguity or unnecessary inference from the part of the listener and they try to expand as much as possible the expressive power of the language, while still minimising its complexity. They thus negotiate shared conventions as part of ongoing interactions [6]. The interactions between the robotic agents used in this experiment take the form of language games in which the agents describe to each other a certain part of the world. The world consists of dynamically moving objects. For example, a hand grasping a block, or a ball rolling against another ball (see figure 1). After a visual analysis of the scene, the speaker selects and conceptualises part of reality, verbalises the conceptualisation, and transmits this to the hearer. The hearer sees whether the description fits with the scene. If this is not the case, or if the speaker calculates that the utterance is potentially ambiguous, or if there is a new construction used by the speaker which is not known to the hearer, then they modify or expand their grammars to be more successful in the future.

The next section describes first in more detail the target of learning: What is case grammar, specifically what aspects of meaning does it express and how is it done in natural languages. Then the learning and invention mechanisms which were used in the current experiment are presented. The paper concludes with a discussion of results and some prospects for future work.

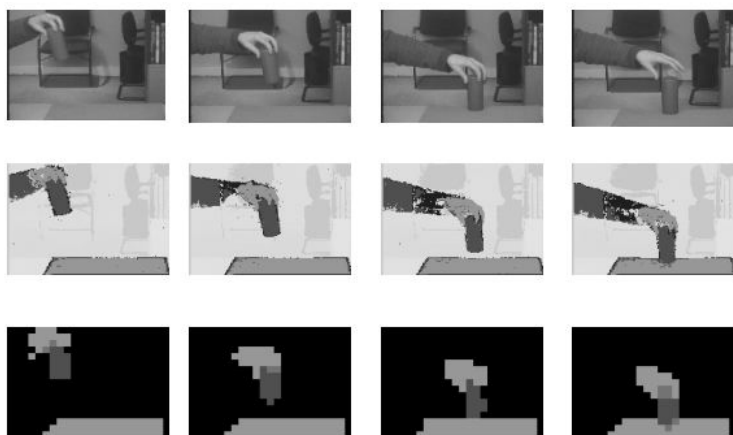


Figure 1: Scene in which a hand is moving towards a block and then grasping it. The top row shows the original scene, the second the scene as processed by the vision system, and the bottom row the abstraction into objects.

2 What is Case Grammar

There is a massive literature on case in linguistics. Even for specific cases, like the dative, one finds a wealth of studies (see for example the articles in [27]). But unfortunately there is very little consensus, resulting in a lot of terminological and theoretical confusion. The main problem is that case manifests itself at different levels and differently depending on the language. Moreover many syntactic phenomena (including case marking) express several aspect of meanings simultaneously, so that it is sometimes difficult to isolate one aspect. Let me start by clarifying the terminology further used in this paper.

2.1 Cases and Case Markers

Fortunately, concerning the external aspects of case, there is a broad consensus. Case is commonly defined as the marking of noun phrases (usually only the noun but it can also be adjectives or articles within the same noun phrase), through the use of affixes, inflections, particles or (prepositional) function words. To restrict the discussion, this paper only focuses on cases with verbs expressing events. Not all languages have extensive case systems. English used to be a strong case-based language (like all Old Indo-European languages) [26], but has basically lost its case system, replacing it mainly by prepositions and only keeping a weak remnant in the pronouns (as in: she vs. her). A well known example of a case-based language is Latin which has different endings for each case: dominus (nominative), domine (vocative), dominum (accusative), domini (genetive), domino (dative and ablative). A distinction must be made between

a *case marker* (e.g. *i* vs. *o* in Latin) and a *case* (e.g. nominative or accusative).

Although there have been many efforts to find universals in case systems, there is now a consensus among typologists that no such system exists, not even within a family of languages [27]. There is enormous variation, ranging from languages with only 2 cases to languages with 40 [2]. Moreover the cases that occur in specific languages are very different, even if occasionally the same terms (like dative) have been used by linguists to describe them. It follows that learning a case grammar implies not only learning the case markers but also learning what cases occur in a specific language. Given the wide variety and the profound evolution in case systems, it is unlikely that the learner has a fixed list of possible cases from which she simply chooses the ones occurring in a specific language.

2.2 Semantic Roles

What do cases mean? Basically there are two systems interfering with each other, each expressing different aspects of meaning:

1. Semantic relations between events and the objects involved. Typically these relations are classified as agent, experiencer, beneficiary, patient, location, time, goal, etc. I will call them semantic *roles* or simply roles. Some authors also use the word case for this [7] or the term thematic role [28]. There is a wide consensus that one of the main functions of cases is to express these semantic roles and that is also the use that will be explored in this paper.
2. Information structures [12] profile the point of view of the speaker for what is given or new, what needs to be emphasised, etc. The information structure influences the selection of grammatical relations (subject vs. direct-object or indirect-object) and this is also expressed occasionally with cases or systems similar to case. Often there is strong interference between the expression of semantic relations (through case) and the expression of grammatical relations [15]. For example, The Japanese particle "wa" indicates the topic of the sentence, whatever the semantic role may be. The only remnant of the case system in contemporary English is now used mainly for expressing grammatical relations. The same subject/verb construction may express different semantic roles as in: The child broke the window (subject = agent). The stone broke the window. (subject = instrument) The car broke an axle. (subject = patient). There are languages which do not have an explicit expression of information structure through grammatical relations but have nevertheless a rich case system [14].

In this paper it is assumed that the expression of semantic roles is the most basic function of case, and only this aspect is explored further.

Just as it has turned out to be impossible to agree on a universal set of cases, it has been equally impossible to find a universal set of semantic roles, even though many linguists, and some AI researchers, such as Schank in his conceptual dependency theory [17] have been trying. On the one hand the roles depend on the categorisation of

events, the ‘agent’ in ”He moved the table” is quite different from the agent in ”He thinks he can do it”. But even more troublesome is that the roles themselves are at least to some extent language specific [25]. It follows that learning a case grammar implies also learning what kind of semantic roles underly a specific language.

2.3 Event-object Relations

This raises the next question, how can these semantic roles be learned? Is there something still more basic? The answer is yes. More basic are the specific events and the relationships they induce on a set of objects. Consider an event where a red object moves towards a green one. The event is ‘move-towards’ and there are two objects involved: a red object and a green one. The red object is the ‘mover’ and the green object the ‘move-target’. Relations like ‘mover’ and ‘move-target’ are specific for the move event. Following standard terminology, I will call these relations the event-object relations. They are the arguments of the predicate characterising the event. Again we should ask the question whether there is a universal set of pre-given predicate-argument structures for describing events or whether every speaker has her own open-ended set that was learned in interaction with the environment perhaps under the influence of language. In this paper, I will assume that the predicates and arguments are given but not that they form a closed set. I believe personally that they are not universally shared (and surely not innate) but this point is not discussed further in this paper.

We conclude that evolving a grammar for case involves (1) evolving a mapping from event-object structures to semantic roles, (2) evolving a mapping from semantic role to cases, and (3) evolving a mapping from cases to case markers. Also all the inverse mappings need to be evolved. The set of language specific case markings, cases, and semantic roles involved in these mappings need to emerge as well.

This task is daunting, particularly because none of these mappings are externally visible to a language learner (the hearer cannot look inside the head of the hearer to see what cases, roles or event-object relations have been used by the speaker) and because mappings are not one to one. The same semantic role can be expressed by many different cases (depending on the linguistic context) and the same case can express many roles, just like words are polysemous and there are many lexical choices for expressing the same meaning. Another difficulty is that languages constantly change. New constructions are invented and spread in the population and so there is no guarantee that different agents all have exactly the same grammar [10].

3 Language games

The analysis in the previous section suggests a scenario for studying the origins and evolution of grammars for case: Starting from a pure lexical languages, three interacting processes must be put to work: (1) the generation of markings for event-object relations, (2) metaphorical and analogical extension of these markings so that semantic roles emerge and hence the markers get a more abstract usage, and (3) evolution

of a surface case system which maps semantic roles onto cases under the influence of syntactic context. In this paper I focus only on the first two steps of this evolution.

The experimental set-up is similar to the one used in our earlier ‘Talking Heads’ experiment [22], that means two pan-tilt cameras looking at real world scenes. But now we use dynamical scenes as opposed to static ones, such as a hand picking up a ball or a puppet called Jack giving a box to another puppet called Jill (see 2). Both agents first look at a scene happening before their cameras. They analyse the image and obtain a stream of event descriptions at several layers of detail. The vision system itself is very complex and described in more detail in [23]. This work is related to other attempts in visual event-recognition such as that by Siskind [?]. It first delineates objects based on colour histograms (see second row in figure 1) and then groups the pixels belonging to the same object together (see bottom row in figure 1). Starting from basic primitive relations like, touching, movement, and appearance, more and more complex descriptions are assembled in real-time by a series of demons which each attempt to recognise in parallel a pattern in the scene. Often this recognition will be unreliable, but saliency, confidence, and hierarchy are used to get a reasonable idea what micro- and macro-events took place.



Figure 2: Scene enacted with puppets so that typical interactions between humans involving agency can be perceived and described.

The filtered result of sensory processing is represented as a series of facts in memory. Next the speaker chooses one event from the set of recent events, conceptualises it, and then verbalises this conceptualisation choosing words for the predicates and applying the constructions of his case grammar to determine cases and case markers. The hearer must lookup the words and decode the case markers, and confront the resulting semantic structure with the facts in his own fact model. The game succeeds if the hearer finds an event in recent memory that is compatible with a description produced by the hearer.

A typical example of output from visual processing for an event in which a red ball

moves away from a smooth green ball is:

```
(move-away-from ev-1) (move-away-from-1 ev-1 obj-1)
(move-away-from-2 ev-1 obj-2) (green obj-2)
(ball obj-2) (hand obj-4)
(smooth obj-2) (ball obj-1) (larger obj-1 obj-2)
(box obj-3) (next-to obj-1 obj-3)
```

There is also a hand and a box in the scene but they do not play a role in the move-away-from event. Facts like these are recorded in the fact memory of each agent. The facts have associated measures of confidence and saliency, as well as a time stamp for beginning and end, and dependencies representing which micro-events were used to assert a macro-event. Facts are continually updated as time goes on. The memory is short term, in the sense that after a certain time period, facts about past events are erased and can no longer be the subject of a game.

Before engaging in a communication, the speaker chooses (randomly) a recent event plus a set of objects related to this event (for example, ev-1 and obj-1 and obj-2). For each object and for the event itself, the speaker then seeks a predicate or conjunctive combination of predicates, that are distinctive for the object or the event. This is similar to the Talking Heads experiment described in earlier papers [22]. Distinctive means that the predicate (or combination) is only valid for the intended object but not for any other object in the scene. Thus ball is not distinctive because both obj-1 and obj-2 are described with it. However green (or smooth) and red are distinctive.

The goal of the hearer is to see whether the event communicated by the speaker is indeed among the recent events. The language game succeeds if the event occurred from the viewpoint of the hearer, and fails if the hearer cannot interpret the utterance in a way that is compatible with events seen in the recent past.

To make the experiments feasible and focus on case, it is assumed that agents already have a lexicon for predicates. Unambiguous English words like RED, or MOVE-TOWARDS are used for the pre-given lexicon so that it is easy for human observers to follow what is going on. This is justified because earlier experiments have shown how populations of distributed agents can arrive at a shared lexicon of words expressing visually grounded categories that identify objects in the world and because we want to focus exclusively in this experiment on case and so everything else is scaffolded.

I now describe three experiments which each reflect one stage in the evolution of the grammar. The first one concerns a purely lexical language without any case marking at all. Agents nevertheless communicate successfully because of the shared situation. In the second experiment, agents invent and learn markers for specific event-object relations. In the third experiment, the agents generalise their markers to develop a layer of semantic roles.

4 A Lexical Language

Even if agents have only a lexicon for the basic predicates, it is already possible to successfully communicate, provided the situation is sufficiently simple and enough information can be drawn from the context to interpret the utterance. This is in line with the point of view advocated by Sperber and Wilson [16], that natural languages are not pure coding systems (where all information is in the message) but inferential coding systems where the sender assumes that the receiver is intelligent enough to figure out what could be meant, based on inference from commonly known facts and the current situation. The message is therefore incomplete and cannot be interpreted without the context. An example is the utterance "move-away-from smooth" (or its equivalent "smooth move-away-from") which does not specify whether it is the smooth object which moves-away or whether there is another object which moves away from the smooth one.

A large number of standard techniques adopted from computational linguistics have been adopted to implement a first experiment where agents communicate with a lexical language without any marking of case. A feature-structure-like representation familiar from unification grammar has been used for the representation of the structures built up during parsing and production. A structure consists of a set of units. They are represented in a parenthesised notation. Each unit carries both syntactic and semantic information. In the interest of clarity, the structures are displayed here separately and relations between them can be inferred because units in both have the same name. An example (taken from the implementation) of a semantic structure is the following:

```
((unit-15
  (goal assertion)
  (subunits
    ((predication unit-16)
     (object unit-17))))
(unit-16
  (goal predication)
  (referent ev-1)
  (meaning
    ((move-away-from ev-1 true)
     (move-away-from-2 ev-1 obj-2)
     (move-away-from-1 ev-1 obj-1))))
(unit-17
  (goal reference)
  (referent obj-2)
  (meaning ((smooth obj-2))))))
```

This structure consists of three units, one for the total, one for the event, and one for one of the objects implied in the event. The semantic information of a unit includes its communicative goal (assertion, predication, reference) and possibly a set of subunits with specific subfunctions. It includes also a referent (extension) which is the grounded

object the unit refers to, and a meaning (intension) which is a (conjunctive) combination of facts which circumscribe the referent. The syntactic information for the same structure is the following:

```
((unit-15 (scope utterance)
  (subunits (unit-16 unit-17)))
 (unit-17 (form (smooth))))
 (unit-16 (form (move-away-from))))
```

Note how the units in the semantic structure reoccur in the syntactic structure. At this point, there is no real syntax, in the sense of word order or marking. The only syntactic information contained in this structure is that unit-17 and unit-16 are represented as belonging to the same utterance.

In this experiment, and all the ones that follow, the same rule formalism COGAR [19] is used, both for the lexicon and the grammar. COGAR uses similar mechanisms as unification grammar and construction grammar. It consists of reversible rules that match and build feature structures through unification. Both sides of the rule feature expressions with variables. [(= X Y ..)] means includes X, Y and the other elements.] Here is an example of such a rule, defining the word "move-away-from":

```
((?unit
  (meaning
    (= (move-away-from ?event ?state)
      (move-away-from-1 ?event ?object)
      (move-away-from-2 ?event ?source)))
    (referent ?event))
  <-->
  ((?unit
    (form (= move-away-from))))))
```

This rule specifies that the word-form "move-away-from" should be used to express an event which is described with the move-away-from predicate. No information is given about how the event-object relations (move-away-from-1 and move-away-from-2 need to be expressed).

In production, lexicon-lookup proceeds by *matching* the left-side of a lexical rule against the semantic feature structures and, if there is a match, *unifying* the right side with the syntactic structure. For example, the left side of the above rule matches with unit-16 in the semantic structure shown earlier (with ?unit bound to unit-16, ?event to ev-1, etc.) and consequently a part of the syntactic structure can be built by unifying the right side with unit-16, thus adding (form (= move-away-from)) to unit-16's syntactic structure, i.e. declaring that unit-16 has move-away-from as one of the elements in its form slot.

In parsing, the right-side is *matched* against the syntactic structure, and, if there is a match, the left side is *unified* with the semantic structure. While unifying, new variable names are introduced for all variables that were not bound. The union of

all the expressions associated with each of the units is the complete meaning of the utterance. Thus given the words "SMOOTH RED MOVE-AWAY-FROM", the final meaning obtained by merging the meanings of the individual units would yield:

```
((move-away-from ?event-1)
 (move-away-from-2 ?event-1 ?source-1)
 (move-away-from-1 ?event-1 ?patient-1) ; from MOVE-AWAY-FROM
 (red ?object-3) ; from RED
 (smooth ?object-4)) ; from SMOOTH
```

To interpret the utterance, the hearer matches this expression against the fact memory, obtaining the following bindings:

```
((?event-1 . ev-1) (?patient-1 . obj-1)
 (?source-1 . obj-2) (?object-3 . obj-1)
 (?object-4 . obj-2))
```

This is indeed among the recent events in the hearer's memory, and so the game succeeds. Note that ?object-3 and ?patient-1 both bind to obj-1 and ?object-4 and ?source-1 to obj-2. In other words, the agent computed that it is the red object which is moving away from the green one and not vice-versa. This information was not expressed verbally but inferred from seeing the scene itself. So a lexical language without marking of event-object relations can lead to successful communication but only because the hearer infers the correct relations from the scene itself and only if there are no other events where the opposite relations hold, which would be the case if we have in the fact memory both

```
(move-away-from ev-1) (move-away-from-1 ev-1 obj-1)
(move-away-from-2 ev-1 obj-2)
```

and

```
(move-away-from ev-2) (move-away-from-1 ev-2 obj-2)
(move-away-from-2 ev-2 obj-1)
```

5 Marking event-object relations.

Obviously, if the object-event relations are expressed explicitly, the communication becomes more accurate and also requires less inference by the hearer. But the question is: how does the speaker know whether additional marking would be helpful? This problem is solved by re-entrance. When the speaker produces an utterance, he first simulates, using his own interpreter and knowledge of the grammar, what the interpretation process will be like for the hearer. This way he can detect possible ambiguities or points where the hearer needs to do more inference than is desirable. Based on this analysis, the speaker can then introduce an expansion of the grammar (in other words the marking of an event-object relation) and reformulate the utterance.

The hearer uses the lexicon to get the basic meaning of the utterance and maps it onto his own fact model, ignoring all markings that he does not know. Usually this leads to a unique interpretation when matched against the fact model. The hearer can try to figure out what the additional marker(s) introduced by the speaker could possibly mean and thus make a very educated guess about unknown rules in the language.

5.1 Invention and Learning Strategy

All this has been implemented in the second experiment, where the agents have the following more sophisticated behavior:

1. The speaker simulates the understanding process of the hearer. In other words, the speaker takes the utterance produced by his own production process and then runs the lexicon lookup and grammatical parsing for this utterance. The results are matched against the fact memory, just as would be the case if he is a listener, and results in a set of bindings. From the bindings equalities can be inferred, i.e. variables which are bound to the same object. For the above example these equalities were:

`?patient-1 = ?object-3 and ?source-2 = ?object-4`

If the utterance already contained information about the event-object relations there would be no equalities. So by computing equalities, the speaker can detect where ambiguity lies, i.e. where the context is needed to disambiguate the utterance, and the speaker can improve the communicative accuracy by marking the event-object relations. In the present implementation, the speaker picks an equality and introduces a marker for the event-object relation involved. For example, to express that the variable `?source-2` is bound to the same object as the variable `?object-4`, a marker (let us say `PO`) can be associated with the word that introduces the binding of `?object-4`, namely `SMOOTH`. The utterance then becomes: `PO SMOOTH MOVE-AWAY-FROM`.

2. If the hearer already knows the marker, in other words has rules to decode it, the interpretation will in a straightforward way lead to the desired set of bindings without equalities. But if the hearer does not know the marker (or the interpretation with the hypothesised meaning of the marker failed), and he is nevertheless able to find a possible interpretation, the hearer can add a new rule to his grammar recording the function of the marker. This is based on the principle that the hearer assumes that there is enough information in the utterance (combined with the context) to uniquely figure out what is meant and if there are unknown elements, these are intended to aid in understanding.

Markers are first introduced with a particular word (like “smooth”) and then carried over to other words, like “green”, if they introduce an object that has the same object-event relation.

5.2 Nature of the Marker Rules

For every event-object relation that needs to be marked (e.g. `move-away-from-2`), three things are done. First a grammatical category is introduced. The category is labeled

'takes-category-X' where X is an identifier. Second, a rule is introduced that maps the event-object relation that needs to be expressed to syntactic information how it should be expressed. This syntactic information contains a declaration that the argument-unit belongs to the grammatical category in question, introduces a new marker-unit which has as associated form "po". There is also an ordering constraint introduced between the marker-unit and the unit which takes the marker, expressed as a constraint (<< ?X ?Y), specifying that the unit bound to ?X needs to follow the unit bound to ?Y. This ordering constraint can be postfix or prefix and is in the example below set to prefix. Here is an example of such a rule which defines a marker for the move-away-from-2 relation:

```
((?predicate
  (meaning
    (== (move-away-from ?evnt)
      (move-away-from-1 ?evnt ?patient)
      (move-away-from-2 ?evnt ?source))))
  (?argument (referent ?source)))
<--->
((?argument (gramcat (== takes-marker-1)))
  (?marker-unit (form (== po)))
  (<< ?marker-unit ?argument))
```

Third, another rule is created by the speaker which declares the word associated with the marked unit as taking this marker:

```
((?unit (gramcat (== takes-marker-1))
  (form (== smooth)))
<--->
((?unit (form (== smooth))))
```

Note that all these rules are bidirectional. In parsing, the matching and unification process ensures that the referent of 'smooth' (when preceded by the marker 'po') becomes the same as the ?source in move-away-from-2. In production, the matching and unification ensures that the right word order is obtained and the appropriate marker inserted.

Here are the steps the hearer goes through when interpreting the utterance "move-away-from po smooth" ("po smooth move-away-from" would give the same result but "smooth po move-away-from" not because po has to precede the noun it qualifies). It starts with the following syntactic structure, in which the ordering of the units is explicitly represented in the ordering-slot of the utterance:

```
((unit-15 (scope utterance)
  (subunits (unit-16 unit-17 unit-18))
  (ordering (unit-18 unit-17 unit-16)))
  (unit-17 (form (smooth)))
  (unit-18 (form (po)))
  (unit-16 (form (move-away-from))))
```

After lexicon lookup and application of the “smooth takes marker-1 rule” we get:

```
((unit-15 (scope utterance)
  (subunits (unit-16 unit-17 unit-18))
  (ordering (unit-18 unit-17 unit-16)))
(unit-17
  (form (smooth))
  (meaning ((smooth ?object1)))
  (referent ?object1)
  (gramcat (takes-marker-1)))
(unit-18 (form (po)))
(unit-16
  (form (move-away-from))
  (meaning
    ((move-away-from ?event ?state)
     (move-away-from-1 ?event ?object)
     (move-away-from-2 ?event ?source)))
  (referent ?event)))
```

and this matches with the right side of the “marker-1 rule” (with ?unit bound to unit-17 and ?marker-unit to unit-18) to yield after unification an equivalence of the ?source and ?object1 variables (renamed as ?source-object1).

```
((unit-15 (scope utterance)
  (subunits (unit-17 unit-18 unit-16))
  (ordering (unit-17 unit-18 unit-16)))
(unit-17
  (form (smooth))
  (meaning ((smooth ?source-object1)))
  (referent ?source-object1)
  (gramcat (takes-marker-1)))
(unit-18
  (form (po)))
(unit-16
  (form (move-away-from))
  (meaning
    ((move-away-from ?event ?state)
     (move-away-from-1 ?event ?object)
     (move-away-from-2 ?event ?source-object1))))))
```

The total meaning of the utterance extracted from this feature structure is:

```
((smooth ?source-object1)
(move-away-from ?event ?state)
(move-away-from-1 ?event ?object)
(move-away-from-2 ?event ?source-object1))
```

When this is matched against the fact memory we get the following bindings:

```
((?event-1 . ev-1) (?source-object1 . obj-2) (?object . obj-1))
```

There are now no more equalities and so inference has been avoided.

5.3 Results

Here are two snapshots of the grammars of an agent who is playing language games about a series of dynamic visual scenes. The output shows the marker, the event-object relation being expressed, the associated grammatical category, the event for which the marker was constructed, and the grammatical rule. Next the words are shown that take the marker, i.e. that belong to the grammatical category associated with the marker.

```
=== agent-2 ===
*** Event-Object Markers ***
ZU: Cause-Touch-3 takes-marker-7 [ev-8, rule-ZU]
BA: Cause-Touch-2 takes-marker-4 [ev-8, rule-BA]
PO: Cause-Touch-1 takes-marker-2 [ev-8, rule-PO]
FI: Cause-Move-Inside-1 takes-marker-1 [ev-7, rule-FI]
*** Words taking markers ***
Ball: ZU           Huge: ZU
Yellow: BA        Her: PO
You: FI
```

After a series of more games, we find:

```
=== agent-2 ===
*** Event-Object Markers ***
DI: Release-2 takes-marker-21 [ev-14, rule-DI]
BO: Release-1 takes-marker-18 [ev-14, rule-BO]
TE: Cause-Halt-1 takes-marker-16 [ev-13, rule-TE]
KI: Cause-Halt-2 takes-marker-15 [ev-13, rule-KI]
LU: Appear-Arg takes-marker-12 [ev-11, rule-LU]
FO: Halt-Arg takes-marker-10 [ev-10, rule-FO]
NE: Move-Arg takes-marker-9 [ev-9, rule-NE]
ZU: Cause-Touch-3 takes-marker-6 [ev-8, rule-ZU]
BA: Cause-Touch-2 takes-marker-5 [ev-8, rule-BA]
PO: Cause-Touch-1 takes-marker-3 [ev-8, rule-PO]
FI: Cause-Move-Inside-1 takes-marker-0 [ev-7, rule-FI]
*** Words taking markers ***
Huge: DI ZU           Striped: DI KI NE
You: BO TE FI        Yellow: KI FO NE BA
Pyramid: KI FO NE    Big: KI NE
It: LU FO            White: LU
Ball: ZU             Her: PO
```

6 Introducing and Marking Roles

Although the agents in the previous experiment indeed progressively eliminate all ambiguity in event-object relations, their grammars do not yet resemble the case systems used in natural language. The population generates a growing number of markers and categories (see figure 3) and is therefore cognitively unrealistic. What is missing is a level in between event-object relations and case-markings, whereby the event-object relations are categorised more abstractly in terms of semantic roles like: agent, patient, source, etc. Case markings are then to be defined in terms of these abstract relations and no longer in terms of the specific event-object relations themselves. The implementation of such a level is easy to do within the formalism introduced earlier. We introduce a new slot called sem-role (semantic role) in the feature structure and a particular event-object relation is categorised with this role. Here is an example of a rule that categorises a particular event-object relation in terms of a role (called role-1):

```
((?predicate
  (meaning
    (== (move-away-from ?evnt)
        (move-away-from-1 ?evnt ?patient)
        (move-away-from-2 ?evnt ?source))))))
<--->
((?predicate
  (meaning
    (== (move-away-from ?evnt)
        (move-away-from-1 ?evnt ?patient)
        (move-away-from-2 ?evnt ?source))))
  (expanded-meaning (== (role-1 ?evnt ?source))))
```

The marking rule now operates over semantic roles:

```
((?predicate
  (expanded-meaning (== (role-1 ?event ?source))))
  (?unit (referent ?source)))
<--->
((?unit (gramcat (== takes-marker-1)))
  (?marker-unit (form (== po)))
  (<< ?marker-unit ?unit))
```

Recall that all rules must be reversible and that the conditional part must match and the concluding part must unify.

6.1 Exploiting Analogy

The key question is where role-abstractions like agent, patient, or source come from. Analogy is proposed to be the key mechanism here. When agents want to explicitly

express an event-object relation, they first try to see whether there is already a marker that expresses an analogous event-object relation. When this is the case, the marker is first generalised to express a new semantic role (if it was not yet a role) and then the new-event-object relation is categorised in terms of this role. If the hearer encounters a marker that was specific to a particular event-object relation in a new situation, he will also use analogy to find the connection between the event used earlier on and the newly encountered event. Note that the speaker forces the analogy on the hearer and the hearer must accept it (and hence learns what kind of analogies the speaker wants to make).

There has been quite a lot of work recently in A.I. to simulate various forms of analogy-making [8]. This is extremely difficult because humans tend to incorporate almost anything in analogical inference and making these inferences is a heuristic as opposed to a rigid logical process. The analogical mapping used in the present experiment is now briefly outlined. More details are found in [24].

The process starts from two events, further called the source-event, for which some relations have already been expressed, and the target event and a target-relation for which marking needs to be constructed. The first step in analogy-making consists in decomposing both events into their primitive micro-events and finding a mapping between them. Here is an example showing how the event-object relation walk-to-1 is mapped onto the relation move-inside-1.

The walk-to event features two event-object relations, walk-to-1 (the agent walking) and walk-to-2 (the target towards which the agent is walking). It consists of four micro-events: The agent does not move, the target does not move, then the agent approaches the target, and then the agent touches the target. This means that the event

```
(WALK-TO-2 ev-100 JILL)
(WALK-TO-1 ev-100 JACK)
(WALK-TO ev-100 TRUE)
```

expands into:

```
(MOVE ev-165641 TRUE) (MOVE-1 ev-165641 JACK)
(MOVE ev-165419 FALSE) (MOVE-1 ev-165419 JILL)
(APPROACH ev-165486 TRUE) (APPROACH-2 ev-165486 JILL)
  (APPROACH-1 ev-165486 JACK)
(TOUCH ev-165633 TRUE) (TOUCH-2 ev-165633 JACK)
  (TOUCH-1 ev-165633 JILL)
```

The move-inside event has also two event-object relations, move-inside-1 (the agent moving) and move-inside-2 (the location in which the agent is moving). It consists of eight micro-events: The agent is visible, the location is visible, the distance between the agent and the location decreases, the location does not move, the agent does not touch the location, then the agent touches the location, and then the agent becomes invisible. The following description of a move-inside event

```
(MOVE-INSIDE ev-163190 TRUE) (MOVE-INSIDE-2 ev-163190 HOUSE-1)
  (MOVE-INSIDE-1 ev-163190 JILL)
```

therefore expands into the following micro-events:

```
(VISIBLE ev-161997 TRUE) (VISIBLE-1 ev-161997 JILL)
(DISTANCE-DECREASING ev-162441 TRUE) (DISTANCE-DECREASING-2 ev-162441 HOUSE-1)
  (DISTANCE-DECREASING-1 ev-162441 JILL)
(MOVE ev-161794 FALSE) (MOVE-1 ev-161794 HOUSE-1)
(TOUCH ev-161801 FALSE) (TOUCH-2 ev-161801 HOUSE-1)
  (TOUCH-1 ev-161801 JILL)
(TOUCH ev-162493 TRUE) (TOUCH-2 ev-162493 HOUSE-1)
  (TOUCH-1 ev-162493 JILL)
(VISIBLE ev-161791 TRUE) (VISIBLE-1 ev-161791 HOUSE-1)
(VISIBLE ev-162665 FALSE) (VISIBLE-1 ev-162665 JILL)
```

Next each micro-event in the target-event is paired with all micro-events in the source-event which use the same predicate. Micro-events which cannot be mapped this way are ignored. The temporal information which is part of the hierarchical event description is not used either. For the mapping from the move-inside event to the walk-to event we get the following result:

```
move-inside event => walk-to event
(TOUCH ev-162689 TRUE) => (TOUCH ev-165633 TRUE)
(TOUCH-1 ev-162689 JACK) => (TOUCH-1 ev-165633 JILL)
(TOUCH-2 ev-162689 HOUSE-1) => (TOUCH-2 ev-165633 JACK)
(TOUCH ev-161796 FALSE) => (TOUCH ev-165633 TRUE)
(TOUCH-1 ev-161796 JACK) => (TOUCH-1 ev-165633 JILL)
(TOUCH-2 ev-161796 HOUSE-1) => (TOUCH-2 ev-165633 JACK)
(MOVE ev-161794 FALSE) => (MOVE ev-165419 FALSE)
                          (MOVE ev-165641 TRUE)
(MOVE-1 ev-161794 HOUSE-1) => (MOVE-1 ev-165419 JILL)
                              (MOVE-1 ev-165641 JACK)
```

A good mapping is defined as one where the filler of the event-object relation of interest (in this case Jack which fills the move-inside-1 role in the move-inside event) always maps onto the same object in the source-event. This is indeed the case here because Jill, which fills the role of move-inside-1 in the walk-to event, always plays the same role in all source micro-events as Jack in the matching target micro-events. Note that walk-to-2 would not extend by analogy to move-inside-2 because the object house-1 maps onto different object roles in the source-event.

Once this analogy has been established, the marker already available for the walk-to-1 event-object relation can be used for marking the walk-to-2 relation.

6.2 Results

Here are again two snapshots of a developing grammar, taken from a multi-agent simulation in which the agents now exploit analogy for building their grammars. In the

first case, the agent has introduced a marker WO for the disappear-arg relation and propagated it to a number of words in the lexicon.

```
=== agent-16 ===
*** Object-event Markers ***
WO: Disappear-Arg takes-marker-57 [ev-30, rule-WO]
*** Words taking markers ***
It: WO           Smooth: WO
Tiny: WO        Blue: WO
Box: WO         Striped: WO
Huge: WO
```

In a later stage, we see that new markers have been constructed but also that the WO marker has been generalised to express a role (role-16) which groups already both the disappear-arg and the cause-move-inside-2 relation.

```
=== agent-16 ===
*** Object-event Markers ***
BO: Cause-Move-Inside-1 takes-marker-60 [ev-32, rule-BO]
KE: Cause-Move-Inside-3 takes-marker-58 [ev-32, rule-KE]
*** Roles ***
role-16: [Disappear-Arg, ev-30]
         [Cause-Move-Inside-2, ev-32]
*** Role Markers ***
WO: role-16 takes-marker-57 [ev-30, rule-mark-role-16]
*** Words taking markers ***
Green: WO           Blue: KE WO
Striped: KE WO     Huge: KE WO
Smooth: WO         Tiny: WO
Me: BO             It: WO
Box: WO
```

The following graph (figure 3) compares the number of markers that have been derived by the agents for two experiments, each using the same series of 1300 language games. The first experiment (top graph) does not use analogy, hence new markers are created for every event-object relation that needs to be expressed. The grammars basically stabilise after about 700 games with 28 markers. In the second experiment (resulting in the two bottom graphs) analogy has been used. There is a graph showing the growth in the number of role markers and another one in the number of event-object markers. Agents generated 4 role markers in the second experiment, covering a wide range of events, and 7 more specific event-object markers, which might still be generalised later. The grammars of the agents stabilise much earlier after 200 games, which proves the point that the use of analogy not only results in a more compact grammar but also that the expressive scope of this grammar is larger.

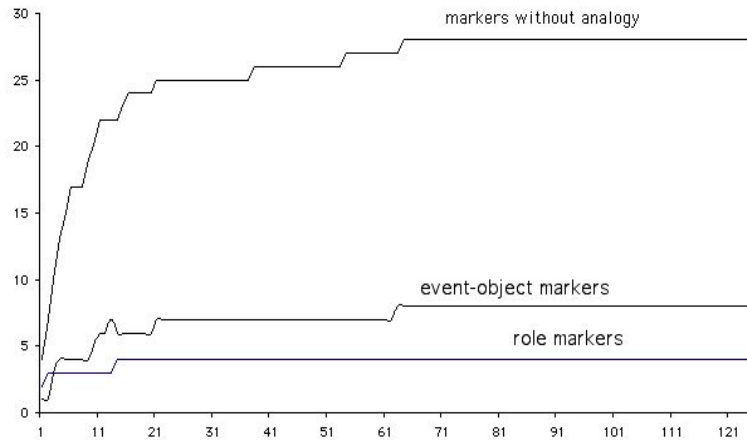


Figure 3: A comparison of two experiments. The top graph shows the number of case markers required for a series of 1300 games, without analogy. The two bottom graphs show both the number of role markers and the number of event-object markers of a more efficient grammar obtained by exploiting analogy.

7 Conclusions and Further Work

This paper discussed a multi-agent simulation in which a population of visually grounded agents develops a natural language like case grammar for marking the relations between an event and participating objects. Starting from a pure lexical language, two stages were discussed: (1) the construction and acquisition of markings for event-object relations, and (2) the metaphorical and analogical extension of these markings towards a more abstract marking of roles. Three very general principles were used. (1) Grammar is seen as meaningful and is introduced whenever the speaker concludes (based on interpreting a self-generated utterance) that the communication could become more precise with additional markings. (2) The language learner uses the situated context to hypothesise possible meanings of unknown markers. The search space of possible meanings for an unknown marker is thus drastically reduced. This avoids the blind inductive inference over a large database of samples typically used in grammar induction [4]. (3) Analogy, which leads to the re-use of existing forms for new meanings, is a fundamental force for introducing new layers in the grammar (such as the layer of semantic roles in between event-object relations and case markings). The same principles would be relevant to evolve grammars for tense, determination, sentence structure, etc.

Although the experiments discussed here show for the first time how the origins of grammar can be modeled by simulating grammaticalisation processes similar to those found in natural languages [9], there is a lot of work that remains to be done, even if we just consider case grammar. Specifically, many (but by no means all) languages

feature an additional layer of surface cases (nominative, accusative, dative, etc.) which comes in between the semantic roles and the case marking. With this additional layer it becomes possible to overload a case marker with additional information about gender, number, etc. or to make the choice how semantic roles map to surface case mappings dependent on syntactic context (like the choice between active and passive sentence). Evolving this additional layer remains a big challenge for the future.

Second, it is known that case markers are not generated by creating a random syllable (as is done in the experiments reported here) but by the re-use of existing words. For example, in many languages case markers have evolved out of verbs in double verb constructions [9] These facts show that there are still many more ‘grammar formation’ mechanisms which need to be formally understood so that they can be incorporated in computer simulations.

Third, we need richer descriptions of events than those which can be extracted from a vision system alone. Human speakers have also active experience with executing the events being talked about (like give, take, move-inside, walk-to, etc.). As shown by recent evidence of mirror neuron systems, the visual recognition of an event is tightly coupled to a (hierarchical) motor control program executing the event, and this would give additional material for analogy-making. For example, the notion of agency is very difficult to detect from vision alone but is very basic when playing a particular role in an event.

Another aspect which needs to be explored is the introduction of population dynamics using populations of more than two agents. This will require that similar mechanisms are incorporated as those we used successfully in earlier lexical experiments [22].

We believe that it will take a large amount of work in detailed computational and robotic modeling to arrive at complex grammars that show many of the crucial features of natural language grammars, but as the simulations reported here show, it is not unrealistic to do so. At the end of such a research program we would have an operational understanding of the forces underlying the invention and acquisition of grammar and a new adaptive language technology in which grammars are not fixed but evolving with the needs of the community.

Acknowledgement

This work was funded in part by a CNRS OHLL grant to the Sony Computer Science Laboratory in Paris and the AI Laboratory of the Free University of Brussels (VUB), as well as by an ESF OHMM grant. I am indebted to Nicolas Neubauer who implemented the unification/matching scheme used for the experiments and Jean-Christoph Baillie who implemented the visual event recognition module.

References

- [1] Anderson, J. (1971) *The Grammar of Case. Towards a Localistic Theory.* Cambridge University Press, Cambridge.
- [2] Blake, B.J. (1994) *Case.* Cambridge University Press, Cambridge.
- [3] Briscoe Edward J. (ed.) (2002) : *Linguistic Evolution Through Language Acquisition: Formal and Computational Models,* Cambridge, UK, Cambridge University Press.
- [4] Broeder, P. and J. Murre (2000) *Models of Language Acquisition. Inductive and Deductive Approaches.* Oxford University Press, Oxford.
- [5] Cangelosi, A. and D. Parisi (eds.) (2001) *Simulating the Evolution of Language.* Springer-Verlag, Berlin.
- [6] Clark, H. and S. Brennan (1991) *Grounding in communication.* In: Resnick, L. J. Levine and S. Teasley (eds.) *Perspectives on Socially Shared Cognition.* APA Books, Washington. p. 127-149.
- [7] Fillmore, C. (1968) *The Case for Case.* In: Bach, E. and R. Harms (eds.) (1968) *Universals in Linguistic Theory.* Holt, Rinehart and Winston, New York. p. 1-88.
- [8] Gentner, D. et.al. (2001) *The Analogical Mind. Perspectives from Cognitive Science.* The MIT Press, Cambridge Ma.
- [9] Heine, (1997) *The cognitive foundations of grammar.* Oxford University Press, Oxford.
- [10] Hopper, (1991) *Emergent Grammar.* In: Traugott, E. and Heine, B. (1991) *Approaches to Grammaticalization. Volume I and II.* John Benjamins Publishing Company, Amsterdam, 1991.
- [11] Kirby, S. and J. Hurford (2001) *The Emergence of Linguistic Structure: An Overview of the Iterated Learning Model.* In: [5].
- [12] Lambrecht, K. (1996) *Information structure and sentence form. Topic, focus and the mental representations of discourse referents.* Cambridge Univ. Press, Cambridge.
- [13] Langacker, R. (1991) *Foundations of Cognitive Grammar. 2 vols.* Stanford University Press, Stanford.
- [14] Mithun, M. (1991) *The role of motivation in the emergence of grammatical categories: the grammaticization of subjects.* In: Traugott, E. and B. Heine (1991) *Approaches to Grammaticalization Vol 2.* John Benjamins, Amsterdam. p. 159-185.

- [15] Palmer, F.R. (1994) *Grammatical Roles and Relations*. Cambridge University Press, Cambridge.
- [16] Sperber, D. and D. Wilson (1986), *Relevance: Communication and Cognition*. Harvard University Press, Cambridge Ma.
- [17] Schank, R. (1973) Identification of conceptualizations underlying natural language. In *Computer Models of Thought and Language*, Schank, R. and Colby, K., (eds.) W. H. Freeman Co., San Francisco.
- [18] Siskind, J. (2000) Visual Event Classification Through Force Dynamics. AAI Conference 2000. AAAI Press, Anaheim Ca. pp. 159-155.
- [19] Steels, L. (2001) Language Games for Autonomous Robots. *IEEE Intelligent systems*, September/October 2001, p. 16-22.
- [20] Steels, L. (2001) Social Learning and Verbal Communication with Humanoid Robots. In: *Proceedings of the IEEE-RAS International Conference on Humanoid Robots*. Waseda University, Tokyo. pp. 335-342.
- [21] Steels, L. (2001) COGAR. A formalism for Cognitive Grammar. Technical Report 2001-17 SONY CSL-Paris.
- [22] Steels, L., F. Kaplan, A McIntyre and J. Van Looveren (2001) Crucial factors in the origins of word-meaning. In Wray, A., editor, *The Transition to Language*, Oxford University Press. Oxford, UK, 2002.
- [23] Steels, L. and J-C. Baillie (2003) Shared Grounding of Event Descriptions by Autonomous Robots. *Journal of Robotics and Autonomous Systems*, 2003.
- [24] Steels, L. (2003) Exploiting situatedness and analogy for learning the meaning of words and grammatical constructions. Submitted for publication.
- [25] Talmy, L. (2000) *Toward a Cognitive Semantics: Concept Structuring Systems (Language, Speech, and Communication)* The MIT Press, Cambridge Ma.
- [26] Van Kemenade, A (1987) *Syntactic Case and Morphological Case in the History of English*. Forist Publications, Dordrecht.
- [27] Van Langendonck, W. and W. Van Belle (eds) (1998) *The Dative*. Volume 1. Descriptive Studies. Volume 2. Theoretical and Contrastive Studies. John Benjamins, Amsterdam.
- [28] Wilkins, W. (1998) Thematic Relations. Vol 21. of *Syntax and Semantics*. Academic Press, Inc. New York.